图书情报工作
第 65 卷 第 3 期　2021 年 2 月

不同学科分类方案下不同学科标准化方法效果的比较研究*

■ 任元秋[1]　王兴[1]　郑钦钦[2]

[1] 山西财经大学信息学院　太原 030006　[2] 优赛思管理系统应用咨询有限公司　上海 200120

摘　要：[目的/意义] 探究不同学科分类方案对于学科标准化方法效果的影响,并比较不同学科标准化方法的效果。[方法/过程] 在 Web of Science 学科分类方案下就比均值法、比中位数法、Z-score 法这三种常用的标准化方法的效果进行比较研究;变更不同粒度的学科分类方案,对这三种标准化方法在 Essential Science Indicators(ESI)、经济合作与发展组织(OECD)学科分类方案下的敏感性进行实证检验。[结果/结论] 结果显示,使用不同学科分类方案并未对各标准化方法的效果产生较大影响,各标准化方法的效果基本保持不变。从 CCDF 引文分布曲线的图形上来看,使用三种标准化方法处理后的 CCDF 曲线形状较原始引文的 CCDF 曲线形状明显更加聚拢,并且三种标准化方法在更换不同粒度的学科分类方案后引文分布情况仍大致相同。结合 top z% 法从定量数值的角度再次进行检验,可以发现,三种标准化方法的效果在变更不同粒度的学科分类方案后基本保持不变,并呈现出如下规律:在截取全局 top30% 以下论文时,比均值法、Z-score 法的标准化效果虽然略有不同,但是都优于比中位数法;截取 top30% –40% 阶段论文时,Z-score 法的优势较为突出;截取 top40% 以上论文时,比中位数法则呈现出明显优于其他两者的效果。

关键词：学科分类方案　学科标准化　标准化效果　引文分布
分类号：G250
DOI：10.13266/j.issn.0252 –3116.2021.03.011

1 引言

论文被引次数是测度科研影响力的重要指标。但由于学科领域之间存在引用习惯以及演化规律等方面的差异,各学科论文的被引次数是不能直接进行比较的[1]。例如生物医学领域,论文更新速度快,发文量较多,论文的被引次数较高;而像数学、地质学等学科,科研周期较长,需要累积达到被引次数峰值的时间也更长,论文被引次数也普遍低于生物医学学科中的论文。为了消除学科之间的差异,对不同学科领域论文的影响力进行公平合理的测度和评价,科研工作者们通常需要使用学科标准化方法对不同学科论文的被引次数进行数学变换,使得标准化后的被引次数,即标准分达到近似同分布的效果,以实现不同学科领域的跨学科比较[2]。

目前越来越多的学者加入到学科标准化方法的研究中,并且基于不同的思想和理论提出了多种标准化方法,如比均值法[3-5]、比中位数法[6]、Z-score 法[7]、百分位数方法[8]、施引方标准化方法[9-10]、逆向工程法[11]、基于交换律的标准化方法[12]等。但是这些标准化方法的提出多是基于学者们的主观经验和直觉,距离实现各学科标准化后引文分布近似同分布的理想状态仍有一定差距。因此,关于标准化方法效果的度量和比较一直是该领域研究的热点。一些学者认为,由于学科引文呈偏态分布,中位数相比于平均值更适合表征数据的集中趋势,认为比中位数法的标准化效果更优于比均值法[13-15]。L. Bornman 等[8,16]认为应该使用分布位置来描述而不是仅用简单的参数指标来评测,并提出将百分位数法用于科研影响力评价效果会更好。张志辉等[17]也就比均值法和 Z-score 法的标准化效果进行了比较,认为 Z-score 法在引文分布的尾部表现出了相对于比均值法的优势,但在引文分布的其他大部分,尤其是低被引部分,效果不及比均值法,并提出效果最优的线性标准化方法。然而,究竟哪种标准化方法更好,更能科学合理地服务于科研评价,目前

* 本文系教育部人文社会科学研究青年基金项目"世界一流大学国际学术话语权研究:国际学术期刊编委的视角"(项目编号:17YJCZH179)研究成果之一。

作者简介：任元秋(ORCID:0000 –0002 –5789 –0361),硕士研究生;王兴(ORCID:0000 –0001 –8176 –5416),副教授,硕士生导师,通讯作者,E-mail:wangxing@sjtu.edu.cn;郑钦钦(ORCID:0000 –0002 –9022 –0589),数据分析师。

收稿日期:2020 –07 –28　修回日期:2020 –10 –25　本文起止页码:84 –92　本文责任编辑:易飞

学界依然没有定论。

此外,标准化方法的效果还受到许多潜在因素的影响,其中,标准化方法效果的实现需要放到具体的学科分类方案下进行。而学科分类本身就是一个较为复杂的问题,不同的国家或地区,甚至是数据库都存在不同的学科分类体系[18]。荷兰莱顿大学的科学与技术研究中心(Centre for Science and Technology Studies, CWTS)提出的皇冠指标(Crown Indicator)采用JCR(Journal Citation Report)主题类作为论文被引次数标准化的参照标准[19]。科睿唯安构建的基本科学指标数据库(Essential Science Indicators, ESI)划分了22个学科,并从国家、机构、期刊、作者等多个层面对文献进行统计分析和排序[20]。这些学科分类方案以期刊为基础对论文进行划分,相应的分类系统将不同期刊划分到了多个学科,仍存在一些不合理之处。以期刊 *Computers & Mathematics with Applications*(CMA)为例, CMA是根据Web of Science学科分类方案收录的"Mathematics, Applied"和"Computer Science, Interdisciplinary Applications"的交叉学科期刊,但在ESI学科分类方案中,却被划分到"Mathematics"这一研究领域[21]。不仅如此,L. Leydesdorff和L. Bornmann[22]还曾指出,Web of Science学科分类方案最初是为了信息检索而不是引用分析开发的,并非直接面向科研评价,各个学科类别之间存在大量的交叉重叠,对于跨学科期刊的分类问题,并没有得到较好的处理。

为此,一些学者甚至自建了学科分类系统,J. Ruiz-Castillo和L. Waltman[23]建立了一个包含5 119个研究领域的论文级分类系统。C. Colliander等[24-25]提出名为"Item-oriented Approach"的方法,不使用任何集群,直接从文章的标题和摘要中提取出名词和形容词将其词根化,并与基于算法构建的聚类分类系统的性能进行了比较。

然而,在面向实际的科研评价过程中,一篇论文由于学科分类方案不同可能被归入不同的学科之下,这对于标准化方法的效果又会产生怎样的影响?例如,一些学者就比均值法的效果对不同学科分类方案的敏感性进行了实证检验:M. Zitt等[26]探究了不同粒度水平的学科领域对比均值法的影响,结果表明,比均值法对不同粒度水平的学科领域是敏感的。J. Adams等[27]也以英国各研究机构为基础验证了M. Zitt的结论。W. Glänzel等[28]分析了676所欧洲大学和研究机构的大量论文,确定了学科标准化引文指标计算的最佳领域深度,提出60个领域的学科分类更适用于机构的研究绩效评估。A. Perianes-Rodriguez和J. Ruiz-Castillo[29]更是指出,在对不同粒度级别的学科分类系统进行选择比较时,使用更高粒度层级的分类系统,通常能够表现出更好的标准化性能。但是,这部分的研究也只涉及比均值法一种学科标准化方法,缺乏对于其他标准化方法更系统的比较。

基于此,本研究将围绕以下两个问题进行讨论:①在最为常用的Web of Science学科分类方案下系统地比较比均值法、比中位数法、Z-score法三种标准化方法的效果,总结三者的引文分布特点以及标准化效果的优劣;②变更不同粒度的学科分类方案,对这三种标准化方法在ESI、经济合作与发展组织(OECD)学科分类方案下的敏感性进行检验,以探究不同学科分类方案对不同标准化效果的影响。这不仅有利于确保标准化效果比较的公平性和准确性,促进学科标准化方法的进一步成熟和完善,同时也将丰富学科标准化的研究内容,完善标准化理论研究体系。尤其是当前"双一流"建设背景下,选择合适的学科分类方案和标准化方法是开展有效科研评价工作的基础和前提,通过探究不同标准化方法和学科分类方案对于评价结果改进的影响,可以更加客观地考察我国"双一流"大学的科研影响力及其同世界一流大学之间的差距,为我国"双一流"大学的建设及科研事业的发展提供有力的支持。

2 数据与方法

本研究将在InCites数据库下分别获取Web of Science、ESI、OECD三种学科分类方案的引文数据,使用比均值法、比中位数法、Z-score法对数据进行标准化处理,通过比较三种标准化方法在不同学科分类方案下的标准化效果,了解学科分类方案与标准化效果之间的联系,分析产生不同标准化效果的原因,并总结使用学科分类方案和标准化方法的规律与准则。研究设计思路见图1。

2.1 学科分类方案的选取

本研究将选取Web of Science、ESI、OECD三种学科分类方案,之所以选择这三种分类方案主要是基于如下考虑:①这些学科分类方案在科研评价中都有着较为广泛的应用,并且它们底层都来源于同一套被引系统,都涵盖在科睿唯安公司的InCites数据库中,这样就保证了它们在数据源上是可比的,避免了不同被引系统之间的差异(如谷歌学术、Scopus与Web of Science三者背后被引系统的差异)对标准化结果的影响[30];②这三种分类方案涉及不同的学科分类粒度,数据下载时,Web of Science将学科划分为255个大类,

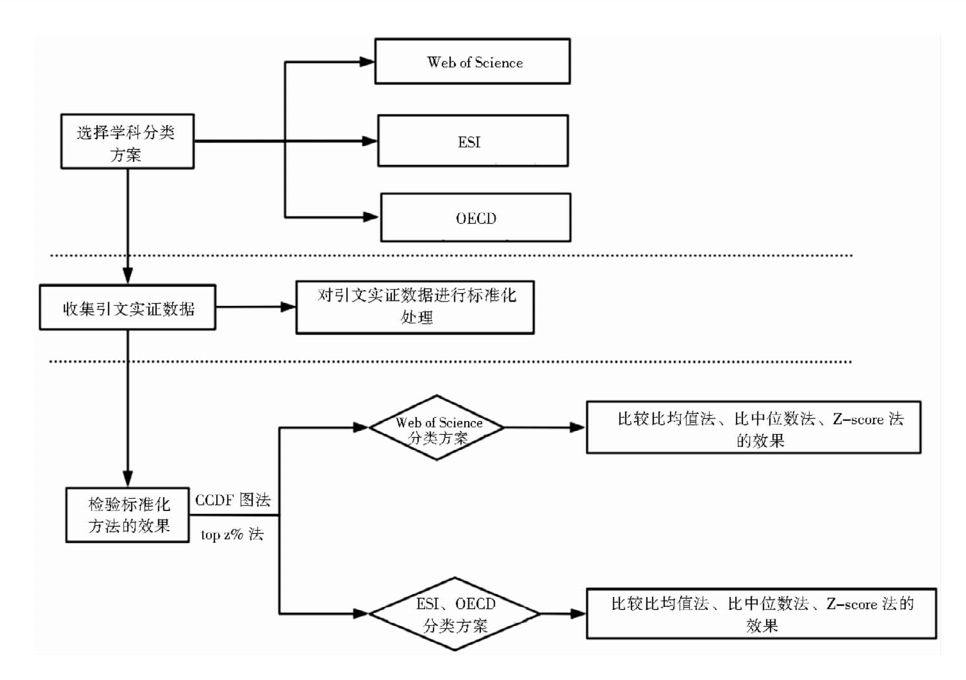


图 1 研究设计思路

OECD 为 45 个,ESI 为 22 个,具有一定的区分度。

2.2 引文实证数据的获取

在 InCites 数据库中分别按照上述三种学科分类方案下载 2013 年文献类型为 Article 的全部学科的论文。其中,Web of Science 中共下载 1 495 337 篇,ESI 中下载 1 362 619 篇,OECD 中下载 1 495 258 篇。论文引文时间窗口为 2013 – 2019 年,保证了 6 年的引文累积时间,使得引文稳定可靠,同时也消除了短时间窗口可能产生的弊端[5,31]。

2.3 学科标准化方法

选取对比的学科标准化方法包括比均值法、比中位数法、Z-score 法。这三种方法是目前较常被使用的具有一定代表性的标准化方法,并且基于现有的引文数据库也都是可操作的、可实现的。在面向实际的科研评价活动时具有一定的可操作性。

具体来说,比均值法,即论文的标准分是原始被引次数与同一学科论文被引次数平均值的比值,其标准分计算公式为:

$$m = \frac{c}{\mu} \qquad \text{公式(1)}$$

其中,m 代表比均值法下的标准分,c 为一篇论文原始被引次数,μ 为该论文所在学科论文的平均被引次数。而比中位数法与比均值法类似,论文的标准分为原始被引次数与同一学科论文被引次数中位数的比值。

Z-score 法则是同时利用集中趋势和离散趋势两种参数的标准化方法,标准分的计算公式为:

$$z = \frac{c - \mu}{\sigma} \qquad \text{公式(2)}$$

同样的,c 表示原始被引次数,μ 表示均值,而 σ 为该论文所在学科引文分布的标准差。

2.4 标准化效果的检验

目前学界判断标准化方法效果优劣的主要依据是考察标准化后不同学科的引文分布情况是否近似同分布,常用的方法为 CCDF 引文分布图法和 top z% 法,有相当数量的文献采用了这两种方法[2,7,32-36]。CCDF 引文分布图法是观察不同学科领域论文的标准分的分布曲线,分布曲线越重合,标准化效果越好;top z% 法则是将所有学科论文的标准分从高到低排序,分别截取全局前 z% 的论文并将这部分论文划分到各个学科,然后计算每个学科落入全局前 z% 的论文占该学科论文总数的实际比例,比例分布越均匀,与期望值 z% 的差距越小,标准化效果越好。这两种方法可以分别从图形以及定量数值的角度对上述三种学科分类方案下不同标准化方法的效果进行检验,且结果较为准确可靠。

其中,某篇论文被引次数的 CCDF 值为:该论文所在引文分布中被引次数大于等于该论文被引次数值的论文占学科论文总数的比例。现用一个例子说明某学科 CCDF 引文分布曲线的绘制过程,假设学科 M 的论文被引次数如表 1 所示,将不同被引次数和对应的 CCDF 值组成的坐标点(0,1.0),(1,0.4),(2,0.1)用线段顺序相连,那么该学科的 CCDF 引文分布情况就是不同被引次数处的 CCDF 值相连而成的曲线,见图 2。

表 1　学科 M 的引文分布情况

被引次数/次	论文数/篇	CCDF 值
0	60	1.0
1	30	0.4
2	10	0.1

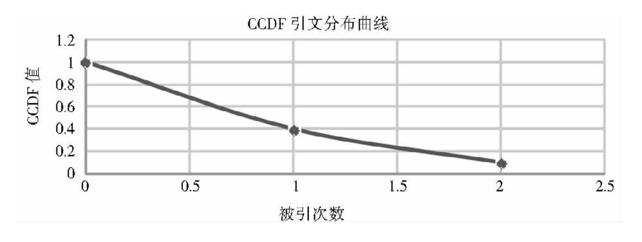


图 2　学科 M 的 CCDF 引文分布曲线

top z% 法使用的指标为论文比例的平均偏差（mean deviation,MD），即各个学科隶属于全局 top z% 的实际论文比例与期望比例 z% 之差的绝对值的平均值，计算公式为：

$$MD = \frac{1}{n}\sum_{i=1}^{n}|p_i - p| \qquad 公式(3)$$

p_i 表示第 i 个学科的实际论文比例，p 表示各个学科期望的论文比例 z%，n 表示学科的总数。当 MD 值越小时，论文的实际比例越接近期望比例，标准化的效果越明显，而 MD＝0 时，则是最理想的状态。

2.5　统计分析

对收集到的三种学科分类方案下的论文被引次数进行整理汇总，计算平均值、中位数、标准差等各个指标，并通过 Python 绘制引文分布图形以检验标准化方法的效果。

3　结果与讨论

3.1　Web of Science 学科分类方案下三种标准化方法效果的比较

使用 Python 对数据进行处理，并绘制 Web of Science 学科分类方案下的 CCDF 引文分布图，结果如图 3 所示。其中，X 轴为论文的原始被引次数或标准分，Y 轴为 CCDF 值。由结果可以发现，在 Web of Science 学科分类方案下，标准化后的引文分布曲线形状较原始引文分布曲线形状更加聚拢，虽然三者仍存在分散的阶段，没有实现所有学科完全理想化的同分布状态，但总体来说，标准化效果都较为明显。尤其是在尾部的高被引部分，都表现出了较高的重合度，与此同时，比中位数法在 CCDF 值 0.5 以上低被引部分，以及 Z-score 法在坐标（0,0.4）对应的“腰部”效果也较为突出，具有明显的同分布趋势。

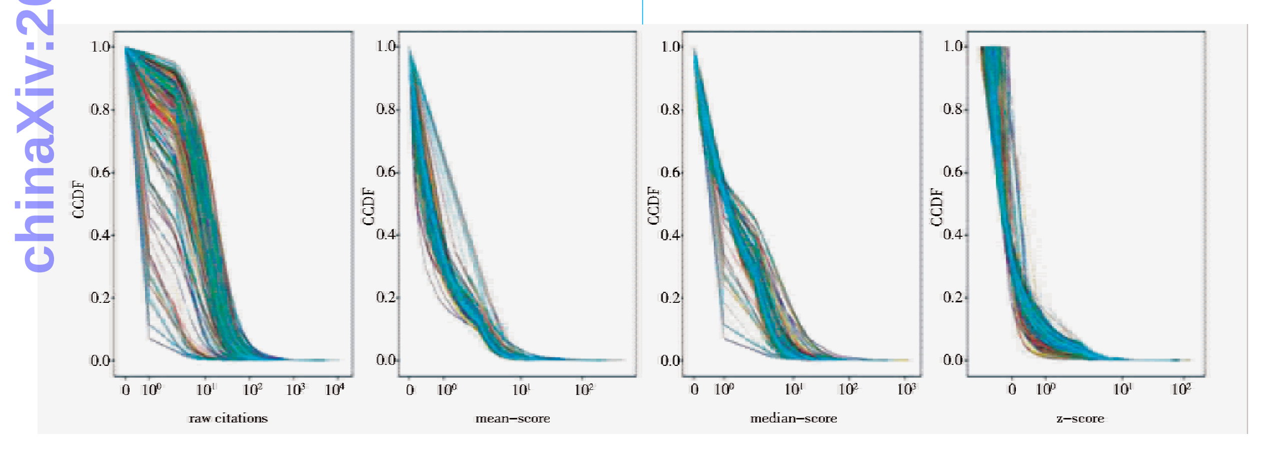


图 3　Web of Science 学科分类方案下不同学科标准化方法的 CCDF 曲线

为了能够细致地比较每个阶段不同学科标准化方法的效果，本文采用更加定量的 top z% 法再次进行检验。分别在 z＝1%、2%、5%、10%、20%、30%、40%、50%、60%、70%、80%、90% 共 12 个截面处，比较每个学科落入全局 top z% 的实际论文比例与期望比例之间的平均偏差。针对性地考察三种标准化方法在高被引部分的细微差别，同时，也对占据绝大多数的非高被引部分的论文进行数值化的比较，尤其 CCDF 引文分布图中三种标准化方法分布的特殊阶段，同样将进行进一步的验证。结果见表 2。

由表 2 可知，当截取全局 top5% 以下的论文时，Z-score 法表现出较优的效果。top5% －20% 处，Z-score 法的效果不及比均值法，但是两者仍都优于比中位数法。top30% －40% 阶段，Z-score 法再次呈现出明显的优势，与 CCDF 图形中坐标（0,0.4）“腰部”相呼应。而 top50% 处这种优势消失。与此同时，比中位数法在 top40% 处也有所变化，MD 值开始小于其他两种方法，top50% 时效果更是突出，并且之后一直优于比均

表 2 Web of science 学科分类方案下
不同截面的平均偏差

top z%	原始被引次数	比均值法	比中位数法	Z-score 法
top1%	0.006 625	0.005 084	0.009 245	0.003 305
top2%	0.012 218	0.007 802	0.016 185	0.007 194
top5%	0.026 759	0.011 876	0.028 529	0.017 164
top10%	0.046 990	0.013 556	0.039 707	0.029 939
top20%	0.082 070	0.019 822	0.047 036	0.038 658
top30%	0.111 492	0.031 299	0.044 842	0.030 864
top40%	0.133 836	0.050 651	0.044 333	0.048 912
top50%	0.154 270	0.071 217	0.048 485	0.088 911
top60%	0.166 574	0.089 134	0.056 448	0.130 365
top70%	0.169 726	0.107 047	0.079 881	0.153 335
top80%	0.161 478	0.117 401	0.102 658	0.159 927
top90%	0.129 545	0.120 900	0.115 763	0.125 937

值法和 Z-score 法，这也与 CCDF 引文分布图中比中位数法呈现出的效果基本一致：CCDF 值小于 0.4 时图形较为发散，大约 0.5 时开始具有明显的聚合趋势，体现了比中位数法在低被引部分的显著优势。

综合两种验证方法来看，三种标准化方法在 Web of Science 学科分类方案下的引文分布曲线，虽然较原始引文分布曲线都明显更加聚拢，但是与理想的同分布状态相比，还存在改进空间，并不能直接评判出哪个是最优的标准化方法。在截取全局 top30% 以下论文时，比均值法、Z-score 法都优于比中位数法，尤其是在 5% 以下的高被引部分，Z-score 法的效果更加突出；在 top30% –40% 处，Z-score 法也有着明显的优势，这与 CCDF 图形中观察到的结果相同，此时论文的原始被引次数等于该学科论文被引次数的平均值；而 top40% 以上，比中位数法的效果则一直优于其他两种方法。

3.2 ESI 学科分类方案下三种标准化方法效果的比较

针对 Web of Science 学科分类方案下比均值法、比中位数法、Z-score 法表现出的规律和特点，本研究将变更不同粒度的学科分类方案，分别考察 ESI、OECD 学科分类方案下三种标准化方法又将呈现怎样的效果。

首先，在 ESI 学科分类方案下，绘制原始被引次数比均值法、比中位数法以及 Z-score 法处理的标准分 CCDF 引文分布曲线，如图 4 所示。标准化处理后的引文分布曲线形状较原始引文分布曲线形状更加聚拢，同时，相比于 Web of Science 学科分类方案下的引文分布情况，Z-score 法在坐标(0,0.4)“腰部”的效果更加明显，比中位数法也在 CCDF 值为 0.4 时就出现了重合的趋势。

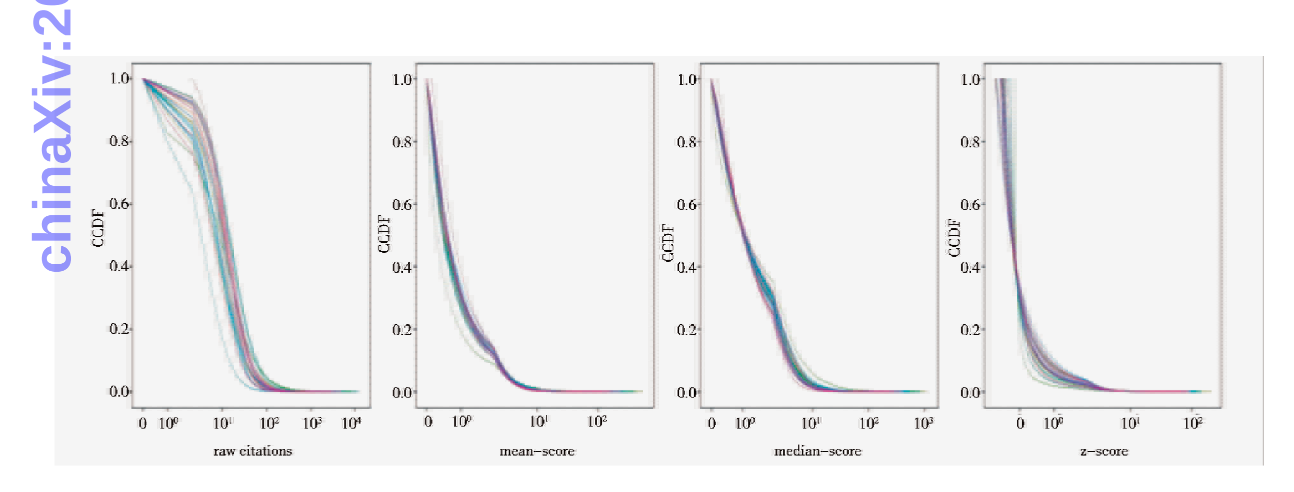


图 4 ESI 学科分类方案下不同学科标准化方法的 CCDF 曲线

为了更加直观地比较三种标准化方法的效果，这里仍使用平均偏差进一步考察每个学科落入全局前 z% 的实际论文比例，见表 3。

可以发现，比均值法在 top30% 以下的 MD 值都较小，效果优于其他两种方法，top1% –2% 时，Z-score 法和比中位数法的 MD 值虽然略大于比均值法，但数值仍都小于 0.01，也有一定效果。top30% 截面处发生变化，比中位数法和 Z-score 法的效果开始优于比均值法，top40% 处更加明显，特别是比中位数法，表现出较优的效果。而 top50% 截面处，Z-score 法又发生变化，效果不如比均值法，与比中位数法更是相差较大，top70% 开始甚至不如未进行标准化处理的结果。三种标准化方法的效果与 Web of Science 学科分类方案下相比，虽然略有差别，但大致相同，并没有因为更换学科分类方案而发生较大变化。

表 3　ESI 学科分类方案下不同截面的平均偏差

top z%	原始被引次数	比均值法	比中位数法	Z-score 法
top1%	0. 006 053	0. 002 153	0. 004 783	0. 002 714
top2%	0. 010 473	0. 003 184	0. 008 043	0. 005 605
top5%	0. 021 482	0. 003 941	0. 014 300	0. 013 725
top10%	0. 035 436	0. 007 545	0. 020 365	0. 022 774
top20%	0. 059 422	0. 014 523	0. 024 050	0. 030 082
top30%	0. 076 733	0. 024 280	0. 020 811	0. 021 554
top40%	0. 093 262	0. 032 641	0. 018 490	0. 026 131
top50%	0. 091 969	0. 040 108	0. 018 393	0. 066 560
top60%	0. 107 347	0. 050 372	0. 021 435	0. 093 210
top70%	0. 093 376	0. 049 521	0. 027 456	0. 108 277
top80%	0. 080 568	0. 048 228	0. 034 384	0. 123 560
top90%	0. 047 038	0. 032 554	0. 029 316	0. 102 174

3.3　OECD 学科分类方案下三种标准化方法效果的比较

最后,对 OECD 学科分类方案下三种标准化方法的效果进行比较,如图 5 所示。同样的,标准化后的引文分布曲线形状较原始引文分布曲线形状都有了明显的聚拢,并且与 Web of Science、ESI 学科分类方案下的标准化效果类似。

为了验证结果的可靠性,也使用 top z% 法对每个截面处的实际情况进行考察,结果见表 4。

由表 4 可知,OECD 学科分类方案下三种标准化方法呈现出了与上述一致的结果,Z-score 法在 top1% 高被引部分以及 top30% －40% 阶段都表现出较好的标准化效果。比中位数法在 top40% 截面开始也表现出与 Web of Science、ESI 学科分类方案下同样的优势。OECD 学科分类方案下,三种标准化方法的效果依然没有发生较大改变。

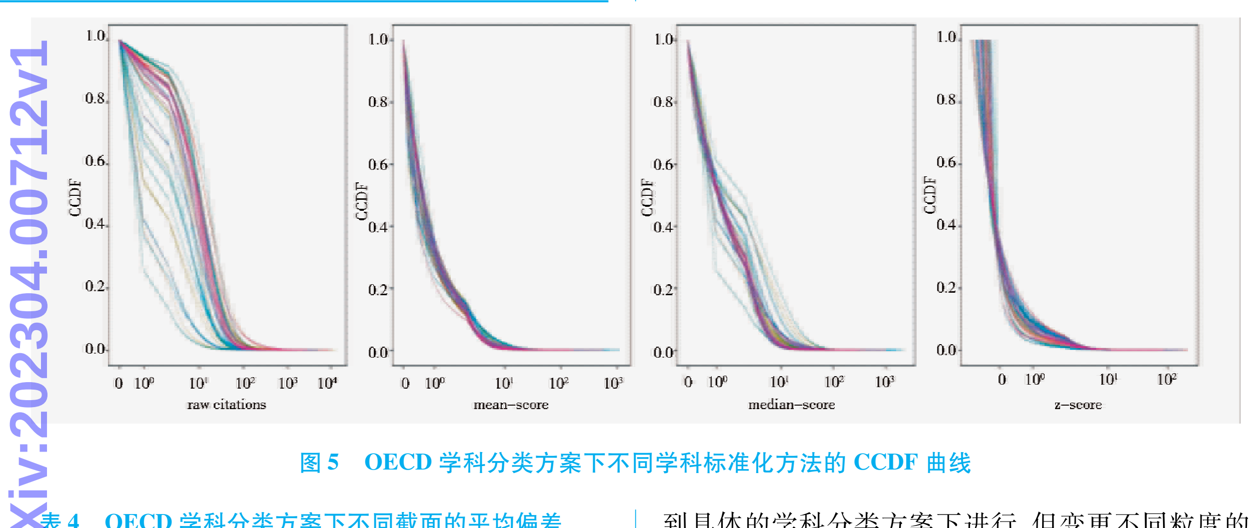


图 5　OECD 学科分类方案下不同学科标准化方法的 CCDF 曲线

表 4　OECD 学科分类方案下不同截面的平均偏差

top z%	原始被引次数	比均值法	比中位数法	Z-score 法
top1%	0. 007 010	0. 005 227	0. 011 737	0. 004 171
top2%	0. 013 123	0. 007 926	0. 019 775	0. 008 639
top5%	0. 029 339	0. 010 181	0. 034 727	0. 019 573
top10%	0. 052 133	0. 012 741	0. 042 124	0. 031 134
top20%	0. 090 785	0. 017 635	0. 043 493	0. 035 468
top30%	0. 121 137	0. 030 053	0. 042 899	0. 027 583
top40%	0. 147 137	0. 051 954	0. 040 207	0. 037 157
top50%	0. 162 276	0. 068 501	0. 044 314	0. 088 848
top60%	0. 170 762	0. 084 892	0. 049 074	0. 136 005
top70%	0. 180 544	0. 104 375	0. 077 000	0. 177 713
top80%	0. 162 170	0. 113 271	0. 106 999	0. 177 997
top90%	0. 146 433	0. 132 716	0. 131 439	0. 142 872

3.4　结果讨论

综合来看,虽然标准化方法效果的实现需要落实到具体的学科分类方案下进行,但变更不同粒度的学科分类方案,各标准化方法的效果仍基本保持不变。从 CCDF 分布图形来看,使用三种标准化方法处理后的 CCDF 曲线形状较原始引文的 CCDF 曲线形状明显更加聚拢,并且三种标准化方法在更换不同粒度的学科分类方案后引文分布情况与变更前大致相同。结合 top z% 法从定量数值的角度再次进行检验,发现三种标准化方法的效果在变更不同粒度的学科分类方案后基本保持不变,并呈现出如下规律:

当截取全局 top30% 以下的论文时,比均值法、Z-score 法虽然略有不同,但都优于比中位数法,表现出较好的效果;截取 top30% －40% 阶段论文时,Z-score 法的效果较为突出;截取 top40% 以上论文时,比中位数法呈现出明显优于其他两者的效果,且一直保持,在 ESI 学科分类方案下,这种较优的效果甚至在 top30% 处就出现了。而比均值法与 Z-score 法在占比较大的低被引论文部分却没有表现出较好的聚拢性,尤其是 Z-score 法,在 ESI、OECD 学科分类方案下部分截面处的效果甚至不及未进行标准化处理时的状态。

呈现上述规律的原因可能非常复杂，或许与标准化方法自身的特点以及学科引文演化规律等有关，也可能是几种因素共同作用的结果。本研究在此做一些尝试性的分析：比均值法是通过论文的原始被引次数除以该论文所在学科的均值来达到引文分布趋于集中的目的的，均值可以反映引文分布的基本特征，却不能反映出引文分布的分散程度以及引文分布位置的重要区别。而中位数是以所有被引次数中所处的位置确定的全部论文被引次数的代表值，一定程度上提高了引文分布数列的代表性，在占据绝大多数的非高被引部分效果比比均值法更好；但在引文累积分布中出现极端变量数值时，中位数由于缺乏一定的敏感性，不受引文分布中高被引的影响，在不同学科论文被引次数差距较大的情况下，不能产生很好的效果来消除这种差异，因此在高被引部分的表现较弱。此外，Z-score 法在 top30% －40% 部分呈现出较优的效果，与此时论文的原始被引次数接近均值有关，说明大多数学科领域中论文被引次数高于均值的论文比例约占 30% －40% 。

在对我国高校科研影响力进行排名时，可以根据上述这些特点和规律，针对不同被引次数水平的论文进行不同的标准化处理。例如，高被引部分可以考虑使用比均值法或 Z-score 法，低被引部分可以使用比中位数法，通过对各单篇论文标准分求和或求平均，以获得大学集合层面整体论文影响力得分，从而有效地识别出表现较好、具有一定科研潜力的院校；科学合理地使用学科标准化指标并对我国高校的科研影响力进行科学的评价，引导我国科学研究事业健康蓬勃地发展。

4 结论

本研究通过绘制 CCDF 引文分布曲线和计算 top z% 平均偏差，分别对 Web of Science、ESI、OECD 三种学科分类方案下不同标准化方法的效果进行了系统的比较，探究了学科分类方案对于学科标准化效果的影响。结果发现，使用不同的学科分类方案并未对标准化效果产生较大影响，在三种具有不同粒度的学科分类方案下标准化方法的效果基本保持不变。从 CCDF 分布图形来看，使用三种标准化方法处理后的 CCDF 曲线形状较原始引文的 CCDF 曲线形状明显更加聚拢，并且三种标准化方法在更换不同粒度的学科分类方案后引文分布情况与变更前大致相同。结合 top z% 法从定量数值的角度再次进行检验，可以发现，三种标准化方法的效果在变更不同粒度的学科分类方案后基本保持不变，且呈现出如下规律：当截取全局 top30% 以下的论文时，比均值法、Z-score 法的效果都优于比中位数法；截取 top30% －40% 阶段论文时，Z-score 法的效果较为突出；截取 top40% 以上论文时，比均值法、Z-score 法的效果不及比中位数法。

本研究也存在一定的局限性：一方面，InCites 数据库下涵盖多种学科分类方案，本研究选取了其中较为常用的且代表不同粒度的三种学科分类方案，今后可继续增加不同粒度的学科分类方案，使得研究更加完善；另一方面，由于非同源论文引用之间存在差异，本研究并未对同样使用率较高的 Scopus、谷歌等学科分类方案进行考察，如何消除这种差异，更加公平合理地衡量不同学科分类方案对标准化方法效果的影响，这也是本研究接下来需要进行扩展和深化的另一个方向。此外，有越来越多的学者加入到学科标准化方法的研究和改进中，并且提出了一些标准化效果较优的新方法，未来可将这些新方法纳入到我们的比较方案中，为我国“双一流”高校科研评价中学科标准化方法及学科分类方案的选择提供更加全面的依据。

致谢：中国知网张志辉博士对本文从整体设计到结果分析再到论文修改均提供了许多宝贵的建议，在此表示由衷的感谢。

参考文献：

[1] 张志辉，程莹，刘念才. 线性学科标准化方法的效果优化及其对科研评价结果的影响——以 39 所“985 工程”大学论文质量排名为例[J]. 情报学报，2015，34(3)：300 －312.

[2] RADICCHI F, FORTUNATO S, CASTELLANO C. Universality of citation distributions: toward an objective measure of scientific impact[J]. Proceedings of the National Academy of Sciences of the United States of America, 2008, 105(45): 17268 －17272.

[3] WALTMAN L, VAN ECK N J, VAN LEEUWEN T N, et al. Towards a new crown indicator: an empirical analysis[J]. Scientometrics, 2011, 87(3): 467 －481.

[4] WALTMAN L, VAN ECK N J, VAN LEEUWEN T N, et al. Towards a new crown indicator: some theoretical considerations[J]. Journal of informetrics, 2011, 5(1): 37 －47.

[5] WANG X, ZHANG Z. Improving the reliability of short-term citation impact indicators by taking into account the correlation between short- and long-term citation impact[J]. Journal of informetrics, 2020, 14(2): 101019.

[6] LEYDESDORFF L, OPTHOF T. Remaining problems with “New Crown Indicator” (MNCS) of the CWTS[J]. Journal of informetrics, 2011, 5(1): 224 －225.

[7] VACCARIO G, MEDO M, WIDER N, et al. Quantifying and suppressing ranking bias in a large citation network[J]. Journal of informetrics, 2017, 11(3): 766 －782.

[8] BORNMANN L. How to analyze percentile impact data meaningful-

ly in bibliometrics: the statistical analysis of distributions, percentile rank classes and top-cited papers[J]. Journal of the American Society for Information Science and Technology, 2013,64(3): 587－595.

[9] LEYDESDORFF L, OPTHOF T. Normalization at the field level: fractional counting of citations[J]. Journal of informetrics, 2010,4(4): 644－646.

[10] ZITT M, SMALL H. Modifying the journal impact factor by fractional citation weighting: the audience factor[J]. Journal of the American Society for Information Science and Technology, 2008,59(11): 1856－1860.

[11] RADICCHI F, CASTELLANO C. A reverse engineering approach to the suppression of citation biases reveals universal properties of citation distributions [J]. PLOS ONE, 2012,7(3): e33833.

[12] CRESPO J A, LI Y, RUIZ-CASTILLO J. The measurement of the effect on citation inequality of differences in citation practices across scientific fields[J]. PLOS ONE, 2013,8(3):e58727.

[13] BORMANN L. Towards an ideal method of measuring research performance: some comments to the Opthof and Leydesdoff (2010) paper[J]. Journal of informetrics, 2010,4(3): 441－443.

[14] BORMANN L, MUTZ R. Further steps towards an ideal method of measuring citation performance: the avoidance of citation (ratio) averages in field-normalization[J]. Journal of informetrics, 2011,5(1): 228－230.

[15] CALVER M C, BRADLEY J S. Should we use the mean citations per paper to summarise a journal's impact or to rank journals in the same field? [J]. Scientometrics, 2009,81(3): 611－615.

[16] BORNMANN L, MUTZ R, NEUHAUS C, et al. Citation counts for research evaluation: standards of good practice for analyzing bibliometric data and presenting and interpreting results[J]. Ethics in science and environmental politics, 2008,8(1): 93－102.

[17] 张志辉.论文影响力的线性学科标准化方法研究[D].上海:上海交通大学,2015.

[18] 陈仕吉,史丽文,左文革.科研机构潜势学科的识别方法与实证分析——以中国农业大学为例[J].情报杂志,2012,31(2): 43－47.

[19] VAN RAAN A F J. The use of bibliometric analysis in research performance assessment and monitoring of interdisciplinary scientific developments[J]. Assessment theory and practice,2003,1(12): 20－29.

[20] 王颖鑫,黄德龙,刘德洪.ESI指标原理及计算[J].图书情报工作, 2006,50(9): 73－75.

[21] HU Z, TIAN W, XU S, et al. Four pitfalls in normalizing citation indicators: an investigation of ESI's selection of highly cited papers[J]. Journal of informetrics, 2018,12(4): 1133－1145.

[22] LEYDESDORFF L, BORNMANN L. The operationalization of "fifields" as WoS subject categories (WCs) in evaluative bibliometrics: the cases of "library and information science" and "science & technology studies"[J]. Journal of the Association for Information Science and Technology, 2016,67(3): 707－714.

[23] RUIZ-CASTILLO J, WALTMAN L. Field-normalized citation impact indicators using algorithmically constructed classification systems of science[J]. Journal of informetrics, 2015,9(1): 102－117.

[24] COLLIANDER C. A novel approach to citation normalization: a similarity-based method for creating reference sets[J]. Journal of the Association for Information Science and Technology, 2015,66(3): 489－500.

[25] COLLIANDER C, AHLGREN P. Comparison of publication-level approaches to ex-post citation normalization[J]. Scientometrics, 2019,120(1): 283－300.

[26] ZITT M, RAMANANA-RAHARY S, BASSECOULARD E. Relativity of citation performance and excellence measures: from cross-field to cross-scale effects of field-normalisation[J]. Scientometrics, 2005,63(2): 373－401.

[27] ADAMS J, GURNEY K, JACKSON L. Calibrating the zoom-a test of Zitt's hypothesis[J]. Scientometrics, 2008,75(1): 81－95.

[28] GLÄNZEL W, THIJS B, SCHUBERT A, et al. Subfield-specific normalized relative indicators and a new generation of relational charts: methodological foundations illustrated on the assessment of institutional research performance[J]. Scientometrics, 2009,78(1): 165－188.

[29] PERIANES-RODRIGUEZ A, RUIZ-CASTILLO J. A comparison of the Web of Science and publication-level classification systems of science[J]. Journal of informetrics, 2017,11(1): 32－45.

[30] BAR-ILAN J. Which h-index? - a comparison of WoS, Scopus and Google Scholar[J]. Scientometrics, 2013,74(2): 257－271.

[31] WANG J. Citation time window choice for research impact evaluation[J]. Scientometrics, 2013,94(3): 851－872.

[32] BORNMANN L, DANIEL H D. Universality of citation distributions-a validation of Radicchi et al.'s relative indicator cf = c/c0 at the micro level using data from chemistry[J]. Journal of the American Society for Information Science and Technology, 2009,60(8): 1664－1670.

[33] RADICCHI F, CASTELLANO C. Rescaling citations of publications in physics[J]. Pysical review e, 2011,83(4): 046116.

[34] WALTMAN L, VAN ECK N J, VAN RAAN A F J. Universality of citation distributions revisited[J]. Journal of the American Society for Information Science and Technology, 2012,63(1): 72－77.

[35] ZHANG Z, CHENG Y, LIU N C. Comparison of the effect of mean-based method and z-score for field normalization of citations at the level of Web of Science subject categories[J]. Scientometrics, 2014,101(3): 1679－1693.

[36] ZHANG Z, CHENG Y, LIU N C. Improving the normalization effect of mean-based method from the perspective of optimization: optimization-based linear methods and their performance[J]. Scientometrics, 2015,102(1): 587－607.

作者贡献说明：
任元秋：数据收集与整理，数据分析，实验设计，论文撰写与修改；
王兴：整体研究思路设计，实验设计，论文指导、撰写与修改；
郑钦钦：数据整理与分析。

Comparison of Field Normalization Effects Based on Different Discipline Classification Schemes

Ren Yuanqiu[1] Wang Xing[1] Zheng Qinqin[2]
[1] School of Information, Shanxi University of Finance & Economics, Taiyuan 030006
[2] Urban Science, Shanghai 200120

Abstract: [Purpose/significance] This paper aims to analyze the impact of different discipline classification schemes on field normalization effects, and compare the field normalization effects of different field normalization methods in different discipline classification schemes. [Method/process] This study focused on two aspects: first, compared the effect of the mean method、median method and Z-score method under the Web of Science classification scheme; second, changed the different discipline classification schemes to test the sensitivity of the three field normalization methods under the Essential Science Indicators (ESI) and Organization for Economic Co-operation and Development (OECD) classification schemes. [Result/conclusion] The results show that the disciplinary classification scheme does not have a significant impact on field normalization effects, and the effects of field normalization methods under different discipline classification schemes remain basically unchanged. Judging from the CCDF distribution graph, the citation distribution after using the three field normalization methods is obviously closer than the original citation count, and the citation distribution of the three field normalization methods after changing the discipline classification scheme with different granularity is still roughly the same. When the top z% method is used to numerically test the field normalization effect, it is found that the effects of the three field normalization methods remain basically unchanged after changing the different discipline classification schemes, and show the following laws: When intercepting papers below 30% of the global top, the Z-score method is slightly different than the mean method, but both are better than the median method; 30% –40% stage, Z-score method shows obvious advantages; more than 40%, the median method is significantly better than the other two methods.

Keywords: discipline classification scheme　field normalization　normalization effect　citation distribution